Usability of parallel processing computers in numerical weather prediction

U.N. Sinha and Ravi S. Nanjundiah

On 23 November 1992, the Department of Science and Technology (DST) convened a meeting to discuss 'Future Supercomputing Strategies for Medium Range Weather Forecasting'. Subsequently it was decided to invite developers of indigenous parallel processing systems (PPS) to evolve suitable strategies of implementation of weather forecasting codes on their respective parallel machines. The aim of this project, as correctly stated by Basu in a recent report in this journal[1], was to demonstrate amongst the scientific community whether the PPS developed in India are capable of handling large applications with reasonable ease and also to benchmark the different PPS machines by running the same application code (namely the spectral model at T80 resolution) with identical initial and boundary conditions provided by a common agency (the NCMRWF). DST realized that India might have a head-start in the field of parallel computing, and its attempt to enhance and augment the indigenous technological base in this (then) emerging field for a well-defined national task was indeed commendable.

Basu was the co-odinator of this exercise and his paper summarizes his findings and views. In the present note, we present certain aspects which appear to have been overlooked by the author and therefore makes his assessment misleading, and offer a different perspective on the project and its international counterparts based on personal experience of one of us (RSN) in India and the US.

Are Indian PPS not good enough?

The title and abstract suggest that the paper is generally about the usability of parallel computing to weather forecasting, while the tone of the paper and its conclusion suggest that Indian PPS are not suitable to meet the requirements of NCMRWF. Basu tries to support this view with the following comments on the Indian exercise:

Poor sustained-to-peak ratio

Basu writes, 'The experience of parallelizing the global spectral forecast model operational at NCMRWF showed that the PPS computers designed and fabricated in India during 1994 could attain a sustained-to-peak performance close to 6%. Since this value is significantly less than the internationally accepted figure, it is possible that the basic design of processor boards used in the machines was not suitable for spectral forecast model.' During the same period as the Indian exercise, Drake *et al.*[2] have published sustained-to-peak ratios for the i860 processor (the processor used in India also by NAL, CDAC and BARC), and we reproduce their tables here. Table 1 displays the performance of the parallel computers in empirical studies, and Table 2 shows the processor's actual performance on meteorological codes.

Considering that the peak speed of i860 is 75 Mflop/s (according to Drake *et al.*[2]), peak of 6% achieved by the Indian PPS was on par with systems elsewhere. Drake *et al.*[2] admit that their experience with the i860 (one of the few processors that have been extensively used in parallel computing applications for meteorology) in regard to its sustained-to-peak speed ratio was less than satisfactory. Therefore it is wrong to conclude that the relatively low value of sustained-to-peak ratio is unique to the Indian PPS (as suggested by Basu). We are not aware on what basis Basu drew his conclusion about 'internationally accepted figures' in 1994.

Scalability

Discussing this issue Basu says: 'To ensure scalability of an application code is not a trivial task even for multitasking, shared memory, vector processing computer. Distribution of data and optimization of inter-processor communication make it even more difficult for a distributed memory PPS.' He further contends, 'Indian machines, however, have not demonstrated scalability clearly and some more effort is

Table 1. Parallel computers used in empirical studies, characterized by operating system version, microprocessor, interconnection network, maximum machine size in experiments (N), message passing startup cost (t_s), per-byte transfer cost (t_b), and achieved per-processor Mflop/s at single and double precision (from ref. 2)

Name	OS	Processor	Network	N
Paragon	SUNMOS 1.6.5	i860XP	16 × 64 mesh	1024
SP2	AIX + MPL	Power 2	multistage crossbar	128

Name	t_s (μs)	t_b (μs)	MB/s (swap)	Mflop/s Single	Mflop/s Double
Paragon	72	0.007	282	11.60	8.5
SP2	70	0.044	45	44.86	53.8

Table 2. Elapsed time per model day and computational rate at T170 resolution on the Paragon and SP2 for double precision and single precision (ref. 2)

			Computational rate	
Name	Nodes	Time/model day (s)	Gflop/s	Mflop/s/node
Double precision				
Paragon	512	1510	1.71	3.3
	1024	814	3.18	3.1
SP2	128	1092	2.27	18.5
Single precision				
Paragon	1024	525.6	4.93	4.8
SP2	64	1606	1.61	25.2
SP2	128	1077	2.40	18.8

required'. Basu is well aware of the fact that a small sequential element in a program can significantly limit the effectiveness of the parallelizing exercise. But the fact that such a small element existed was neither apprehended by the experts at NCMRWF nor by the developers (who, it must be stated, did not have much earlier experience with the T80 code). The NCMRWF T80 global spectral model has its origins in the NCEP model, which has been largely shaped by Sela with Basu as one of the co-authors[3]. Sela's experience in parallelizing this model on a shared memory vector parallel machine (clearly the author's favourite), viz. C90, is very succinctly summarized in Figure 1 (reproduced from Sela[4]).

We would like the reader to note that the efficiency of the C90 with 4 processors was 77.5% and with 8 processors, it was 68.75%. Hence the Indian efforts in the DST project were comparable to efforts elsewhere at the same time (with the disadvantage of little support from the industry in contrast to the close interaction between industry and research groups in most efforts elsewhere).

Basu correctly states that, unlike the implementations of the ECMWF model[5] and the NCAR model[2] where considerable effort was devoted to developing codes that were scalable, Indian PPS developers did not make efforts in this respect. The following must, however, be stated:

1. This was the first implementation of the model, and the general experience is that such first implementations of any software are rarely optimal.
2. The project was closed in March 1996, just as these initial implementations were completed.
3. The efforts of PPS developers after March 1996 have not been considered in Basu's paper, on the pretext that model outputs have not been examined!

Out of scientific curiosity we have conducted further studies on the scalability of this model[6]. We have found that the initial parallel implementation of the NCMRWF code has a sequential component of 4.7% and its scalability on an ideal machine (i.e. with maximum theoretical efficiency, with infinite bandwidth for communication) is presented in Table 3.

Table 3. Comparison of maximum theoretical and actual achieved efficiencies on a 4 processor SGI power challenge (ref. 6)

No. of procs	Maximum theoretical efficiency		Efficiency achieved	
	Case A	Case B	Case A	Case B
1	100.0	100.0	100.0	100.0
2	95.5	99.5	93.3	96.1
4	87.6	98.6	81.4	88.9
8	75.1	96.9	–	–
16	58.5	93.5	–	–
32	40.5	87.5	–	–
64	25.2	82.3	–	–

The cause of poor efficiencies in Sela's or our earlier implementation can now be explained on the basis of this table. Sela's implementation uses the strategy of parallel implementation of grid space computations (Case A in Table 3). However, we have further refined the load decomposition strategy (Case B). This refinement now includes concurrent computing of the linear part of the model, in addition to decomposition of loads in physical space. The sequential part by this strategy reduces to 0.34%, and the scalability consequently improves dramatically. It must be pointed out that in the present version, the computation of the linear part is conducted on the summed coefficients, whereas this could be done on the modes themselves (as modes do not interact in this part of the model).

Had these modifications been performed as part of the Indian project on the PPS, the results would have been less misleading. However, we need to point out that we could arrive at these conclusions and alternate strategies only after the initial parallelizing exercise and after studying the results of this effort. It is disconcerting that Basu (one of the co-authors of the NCEP/NCMRWF model) missed this critical aspect of Sela's parallelization, i.e. the technique of parallelizing computations in physical grid space alone would not

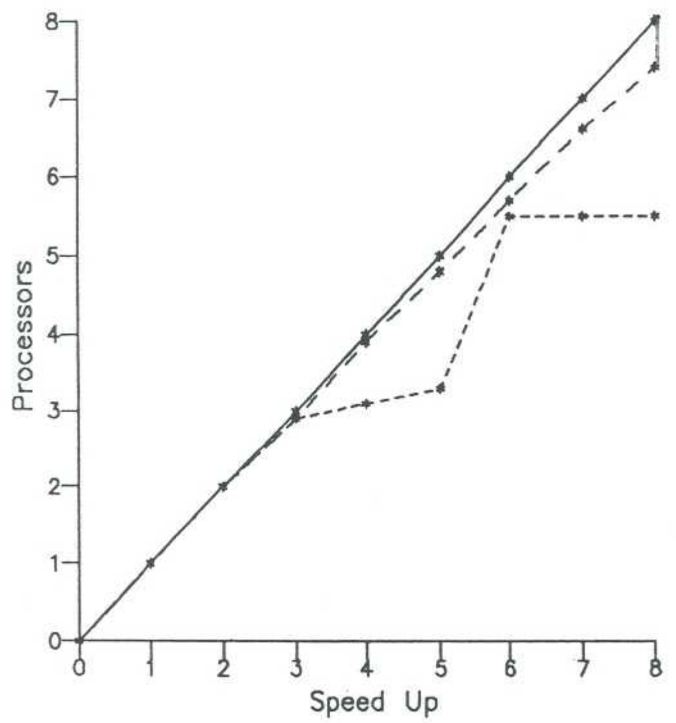

Figure 1. An ATEXPERT report on parallelization on the CRAY C90. The solid line shows ideal speedup, the lower dashed line speedup when the inverse transform is structured to ensure reproducibility, and the upper dashed line speedup when this feature is disabled (from ref. 4).

be scalable (though Sela's study was published during the course of the Indian exercise). Thus, the Indian parallelizing exercise (with its constraints of limited resources and with little support from industry) was actually on par with international efforts at that time.

Reproducibility

During the course of this exercise, reproducibility was considered a major issue. While reproducibility undoubtedly is to be considered, it was given far more importance than was necessary or scientifically justifiable. Thus a significant part of the time was spent on explaining the small systematic differences between the outputs obtained on the PPS and the Cray. However, international experts' views on this issue are far more relaxed. The differences between the results of the IEEE-compliant RISC machines and the non-IEEE compliant Cray were presented at the 7th ECMWF Workshop on the Use of Parallel Machines for Meteorological Applications[7]. The response was that the observed differences could be caused by a problem in a particular segment of the Cray's memory. Even a major modelling centre such as GFDL (the Geophysical Fluid Dynamics Laboratory) has taken a far more lenient view on correctness, reproducibility and validating the parallel implementation[8] than the Indian monitors did. We reproduce here their views on validation:

'Verification of the correctness of the model on a different system is an important step that must be undertaken before the model can be used on that system. Differences in compilers, system libraries, internal number representations, and machine precision can all have significant impact on the answers and make correctness extremely difficult to guarantee. This is particularly true of certain SKYHI (GFDL's GCM) diagnostic sums which involve differences of large numbers and are therefore an effective measure of round-off error. After a single time step, differences between (Cray) YMP and CM-5 (Connection Machine 5, a parallel computer) simulations in the more sensitive diagnostics were less than 10^{-12}, and the largest of these were attributable to roundoff error. After one day, the differences grow to about one per cent for the most sensitive diagnostics and for long simulations only qualitative comparisons can be made.'

In contrast, NCMRWF experts insisted that the results be identical to begin with, and considerable time was spent in convincing them of the correctness of the parallel implementation. Had this time been spent on issues such as scalability, the quality of parallel implementation might have improved further.

Cost effectiveness

In India, unlike in the West, parallel processing has evolved as a strategic necessity and has proved to be extremely cost-effective. The total developmental cost at NAL for parallel processing, over the last decade, is about Rs 2 crores. Basu says, 'Such benefit in the unit cost of computing can more than compensate the large manpower investment required to rewrite the large application code like the forecast model'. It must be mentioned that the budget for CHAMMP (Computer Hardware Advanced Mathematics and Model Physics) initiative of US Department of Energy (DoE) was of the order of a few million dollars, exclusively for the development of a scalable parallel model! The Indian initiative (which included both hardware and software) was conducted at a fraction of this cost.

In the light of Basu's comments and their misleading implications, it is perhaps now necessary that the country should think in terms of an alternate centre dedicated to research on issues related to development of weather forecast models suited/tailored to parallel computers and running them in a semi-operational mode, if the Indian parallel computing initiative for weather forecasting is not to die. If such a centre were to be started *ab initio*, the investment could be around Rs 2 crores (including infrastructure, a parallel computing platform and personnel) over a period of five years. However if establishments having infrastructure and computing platforms, e.g. universities, national laboratories, IITs, IISc, etc. are willing to take up this task, the investment may perhaps be lower for augmentation of the existing facilities.

PPS in India – A global perspective

In case of parallel computing in general and its application to meteorological computing in particular, India had a clear head-start. One of us (RSN) visited Argonne National Laboratory in 1992 and found that our efforts were on par with that of the CHAMMP initiative. Thus DST's decision to examine the feasibility of implementing the global spectral model was a sagacious one. In direct contrast, Basu's 'cautious' approach would lead to importing newer models running on off-the-shelf platforms (which may not be the state-of-the-art machines). This will lead to the perpetuation of the obnoxious 'black box' culture and fritter away, at tremendous cost, all the technical gains made by imaginative use of parallel computing in India. The prospect of an Indian weather forecasting model addressing problems specific to the tropics will recede further if such an attitude continues.

This, interestingly, is in stark contrast to the approach taken by other developing countries such as Brazil. These countries are investing large sums of money (in excess of a million dollars) to develop indigenous parallel computers and weather forecasting models tailored to their needs. Specifically, they are laying great stress on the reverse engineering of existing codes, to gain in-depth knowledge of underlying processes – which hitherto has been exclusive to the developed world.

Even with PPS available in March 1996, operations (five-day) forecasts could be produced about four times a day, and thus could meet the operational requirement. This, however, is not to downplay the computational needs (for research and development) of NCMRWF, but to record the fact that PPS were capable of satisfying the operational requirements of NCMRWF even in March 1996. In retrospect, all we can suggest is that a golden chance to perform weather forecasts on Indian machines using parallel version of forecasting codes implemented by Indians was missed.

Finally, a word about the reliability of Indian PPS. One of these systems was on display and running continuously at Pragati Maidan, New Delhi during the peak of summer in the year 1995, without any air-conditioning, and many top DST officials were witness to this.

In conclusion, we are of the firm view that the kind of 'caution' exhibited in Basu's assessment is precisely the reason why, even when we find ourselves on a position of some scientific or technological advantage internationally, lack of imaginative decision-making or a peculiar technological timidity works to throw away that advantage. Are we going to embrace parallel computing for meteorology only after everybody else in the world has done so – and then rush to buy those systems from elsewhere, having starved our promising programmes by rejection?

1. Basu, B. K., *Carr. Sci.*, 1998, 74, 508-516.
2. Drake, J.. Foster, 1., Michaiakes, Toonen, B. and Worley, P., J. Parallel *Cnnaput.*, 1995, 21, 1571-1591.
3. Kalnay, E., Seta, J.. Campana, K., Basu, B. K., Schwarzkopf. %f.. Long. P., Caplan, M. and Alpert, J-. Documents, tion of the Research Version of the NMC Medium Range Forecast Model. National Meteorological Centre, Washington DC, USA, 1988-
4. Sela, J. G., *J. Parallel Comp* il-, 1995 21,1639-1654.
5. Barrns, S. R. M., Dent, D., Isaksen, L, Robinson, G., Mordzynski, G- and Wol_ lcnwebcr, *F., J. Parallel* Crmipur., 1995, 21,1621-1638.
6. Venkatesh, T. N., Rajalakshmy Sivaramakrishnan, Sarasamma, V. R, and Sinha, U. N.. NAL PD FS 9708, National Aerospace Laboratories, Bangalore, 1997.
7. Sinha, U. N. and Nanjundiah, R. S., Procecdings of 7th ECMWF Workshop on the Use of Parallel Processors in Meteurology (eds Hoffm an, G. R. and Kreitz. N.), World Scientific, Ncl,s Jersey, 1997.
8. Jones, P. W., Kerr, C. L. and Heniler, R. S, *J. Parallel* C anpur., 1995. 21. 1677-1694.

U. N. Sirlha is in Me Flosolver Unit, National Aerospace Laboratories, Kodifzulli Campus, Barzgalore 560 01?, 1 dia; and Ravi S. Nanjundialt is in the Centre for Attnosnheric and Oceanic Sciences, *Indian* Institute *of* Science, 13angalore *560 012, India.*